Amendments to Specification

Please amend the specification as set forth hereafter.

-o-0-o-

Please amend the paragraph starting at page 7, line 1 as follows:

An approach to mapping that increases the number of symbols in the alphabet is a mapping that forms combinations of symbols in the input alphabet. In such an implementation pairs (or triples, quadruples, and so forth) of symbols can be mapped to a single output symbol. For example, suppose the input sequence begins with "ACDF..." (SEQ ID NO:1) . The first two symbols "AC" could map to the output symbol "a". The next pair of symbols "CD" might map to the output symbol "b". The symbols "DF" might map to "c", and so forth. In this specific example groups of n adjacent symbols are mapped to one output symbol. The output sequence will be shortened by (n-1) symbols due to end effects, but the number of possible symbols in the output alphabet is increased. If there are twenty symbols in the input alphabet, then there are a possible 20^n symbols in the output alphabet. Thus, taking n=2, there are a possible four hundred symbols in the output alphabet. With n=3, there would be eight thousand possible symbols in the output alphabet.

-o-0-o-

Please amend the paragraph starting at page 7, line 25 as follows:

The basic implementation of the method of the present invention may be understood by considering the forty-seven place sequence S_1 and the fifty-four place sequence S_2:

S_1:
ECGHHAFSDYQWVDDENPLQKVPTSKPPFTVGDIKKAIPPHCFQRSL
(SEQ ID NO:2)

S_2:
CEVGVVLRKVKPVSKVPIVFQRSLVPTPHVLRKAWVCVYEAGHHQYWF
YGWVNG (SEQ ID NO:3)

-o-0-o-

Please amend the paragraph starting at page 8, line 3 as follows:

The pattern **L.KV........V........PH** (SEQ ID NO:4) is found in both sequences (shown underlined in the above statement of the sequences S_1 and S_2). Here, the dots represent locations where the symbols in the two sequences do not match, and are thus considered placeholder positions in the pattern.

-o-0-o-

Please amend the paragraph starting at page 12, line 16 as follows:

Reading out patterns is now simple. The collection of the symbols tabulated for each value of difference in position (i.e., each row) in the Pattern Map defines a pattern in the first sequence that is repeated in the second sequence. Each row of the Pattern Map is a pattern of symbols contained in sequences S_1, S_2. The pattern, in symbolic form, is determined by consulting S_1 to determine the symbol at the location indicated by the Pattern Map index. For example, Pattern Map row 35 is the above-mentioned pattern L.KV........V........PH (SEQ ID NO:4). The pattern is constructed by noting the relative positions of these symbols and inserting the appropriate number of placeholders (one placeholder between the L and K, eight placeholders between the V and V, and eight placeholders between the V and P).

-o-0-o-

Please amend the paragraph starting at page 13, line 33 as follows:

If, for example, it were desired to identify all those patterns that include more than four symbols, it may be seen by examination of the second column of the Pattern Map that there are eleven patterns of four or more symbols, thus:

Pattern	Row Index
P..FQRSL (SEQ ID NO:5)	Line 24
V.....P.....I............L (SEQ ID NO:6	Line 31
L.KV........V........PH (SEQ ID NO:4)	Line 35
P.SK.P..........P (SEQ ID NO:7)	Line 36
P..KVP.......V (SEQ ID NO:8)	Line 41
V........V.......T.....KA (SEQ ID NO:9)	Line 44
C..............P........P...V (SEQ ID NO:10)	Line 46
E.G...............Q......P........	
.......Q (SEQ ID NO:11)	Line 48

V........VPT...............H (SEQ ID NO:12)   Line 50

L...P.......V...........F (SEQ ID NO:13)   Line 52

E.GHH....Y.WV (SEQ ID NO:14)   Line 86

-o-0-o-

Please amend the paragraph starting at page 15, line 22 as follows:

A second selection criteria that may be used to implement filtering is the "span" of a pattern. By "span" is meant the total number of places from the first symbol of a pattern to the last symbol of a pattern. Thus, the five symbol sequence illustrated in line 36 of the above example:

P.SK.P..........P (SEQ ID NO:7)

has a span of seventeen (i.e., five symbols plus twelve placeholders).

-o-0-o-

Please amend the paragraph starting at page 20, line 10 as follows:

Figure 3 illustrates the method of the present invention in discovering corrupted patterns. This figure shows two eleven-symbol sequences ([DPUTPNOUNDT (SEQ ID NO:15) and DTUPPNOUNOT (SEQ ID NO:16)] written in the alphabet D,N,O,P,T,U. The sequences are shown in sequence order and then repeated in symbol-sorted order to make clear the construction of the Pattern Map in the lower part of the Figure 3. The Pattern Map is written somewhat more elaborately than in Figure 2 in that each row of the Pattern Map is presented in the form of the (position index, total index, symbol) triple. The position index is exactly the index stored in the Pattern Map in Figure 2. The total index is the sum of the position index and the Pattern Map row index (also known as the "difference in position" value). The symbol corresponds to the position index and is included in Figure 3 primarily for clarity of explanation.

-o-0-o-

Please amend the paragraph starting at page 20, line 35 as follows:

The reference pattern is simply described by the set of position indices (1,5,6) where the duplicate occurrence of the value one is ignored. This pattern is "P...NO" (SEQ ID NO:17).

-o-0-o-

Please amend the paragraph starting at page 21, line 4 as follows:

For the first form of the corrupted pattern the symbol "P" having the total index value of fourteen is selected (P:1, 14). The total indices of the symbols in the first form of corrupted pattern are thus (14,19,20) and the resulting pattern is "P....NO" (SEQ ID NO:18) , having an extra insertion between the P and N relative to the reference pattern ("P...NO") (SEQ ID NO:17) .

-o-0-o-

Please amend the paragraph starting at page 21, line 11 as follows:

For the second form of the corrupted pattern the symbol "P" having the total index value of fifteen is selected (P:1,15). The total indices of the symbols in the second form of corrupted pattern are (15,19,20) and the resulting pattern is "P...NO" (SEQ ID NO:17) , which is identical to the reference pattern.

-o-0-o-

Please amend the paragraph starting at page 21, line 17 as follows:

One corrupted pattern relative to the original reference pattern has thus been discovered. The occurrence of the reference pattern in the first sequence is DPUTPNOUNDT (SEQ ID NO:15) and the occurrence of the reference pattern in the second sequence is DTUPPNOUNOT (SEQ ID NO:16) (where the symbols of the pattern are underlined). By definition, the corrupted pattern may not be present in both sequences (otherwise it would have been discovered as a pure pattern). By examining the two sequences it may be seen that the corrupted pattern does not occur in the first sequence, but the corrupted pattern does occur as DTUPPNOUNOT (SEQ ID NO:16) in the second sequence.

-o-0-o-

Please amend the paragraph starting at page 26, line 12 as follows:

As an example, the pattern

L.KV........V........PH (SEQ ID NO:4)

may be represented in FlatMOT table form as:

(L,0), (K,2), (V,3), (V,12), (P,21), (H,22) [[,]] .

-o-0-o-

Please amend the paragraph starting at page 33, line 3 as follows:

The result of tuple-Sort followed by tuple-Squeeze is shown in ~~Figure 7~~
Figures 7A and 7B (labeled "squeezed"). All patterns contained in this tuple-table can
be read. A pattern occurs as a contiguous set of tuple-table entries that share indices in
all but the master column. Thus, for example, the first four rows of the (0,1) squeezed
tuple- table read out as "BCDA" (SEQ ID NO:19) . The next two rows are the pattern
"DA" (SEQ ID NO:20) . The last two rows are "AD" (SEQ ID NO:21) . These
patterns are spaced according to the indices in the primary column. An example of a
somewhat more distributed pattern can be seen in the (0,2) tuple-table, where the
pattern "CD.C" (SEQ ID NO:22) is found. Note that the placeholder between the last
two symbols of the pattern is due to the skip in the primary column from indices four
to six.